

#17



PCT

## RAW SEQUENCE LISTING

DATE: 02/11/2003

PATENT APPLICATION: US/09/830,914C

TIME: 14:29:17

Input Set : A:\pf0621usn\_subseqlist.txt

Output Set: N:\CRF4\02112003\I830914C.raw

C--> 1 <110> APPLICANT: INCYTE GENOMIC, INC.; TANG, Y. Tom;  
 2 CORLEY, Neil C.; GORGONE, Gina A.;  
 3 GUEGLER, Karl J.; BAUGHN, Mariah R.  
 5 <120> TITLE OF INVENTION: MYOSIN HEAVY CHAIN HOMOLOG  
 7 <130> FILE REFERENCE: PF-0621 USN  
 9 <140> CURRENT APPLICATION NUMBER: US 09/830,914C  
 10 <141> CURRENT FILING DATE: 2002-11-06  
 12 <150> PRIOR APPLICATION NUMBER: PCT/US99/26177  
 13 <151> PRIOR FILING DATE: 1999-11-05  
 15 <150> PRIOR APPLICATION NUMBER: US 60/172,248  
 16 <151> PRIOR FILING DATE: 1998-11-05  
 18 <160> NUMBER OF SEQ ID NOS: 4  
 19 <170> SOFTWARE: PERL Program  
 21 <210> SEQ ID NO: 1  
 22 <211> LENGTH: 612  
 23 <212> TYPE: PRT  
 24 <213> ORGANISM: Homo sapiens  
 26 <220> FEATURE:  
 27 <221> NAME/KEY: misc\_feature  
 28 <223> OTHER INFORMATION: Incyte ID No: 1929760CD1  
 30 <400> SEQUENCE: 1  
 31 Met Phe Cys Pro Pro Gln Val Ser Cys Ser Leu Ser Leu Met Pro  
 32 1 5 10 15  
 33 Arg Leu Pro Ser Ile Arg His Trp Gln Gly Pro Ser His Pro Gly  
 34 20 25 30  
 35 Phe Leu Gly Pro Leu Phe Pro Ile Cys Ser Leu Gln Trp Pro His  
 36 35 40 45  
 37 Gly Phe Ser Ala Ile Phe Pro Gly Leu Leu Asp Val Tyr Gly Phe  
 38 50 55 60  
 39 Glu Ser Phe Pro Asp Asn Ser Leu Glu Gln Leu Cys Ile Asn Tyr  
 40 65 70 75  
 41 Ala Asn Glu Lys Leu Gln Gln His Phe Val Ala His Tyr Leu Arg  
 42 80 85 90  
 43 Ala Gln Gln Glu Glu Tyr Ala Val Glu Gly Leu Glu Trp Ser Phe  
 44 95 100 105  
 45 Ile Asn Tyr Gln Asp Asn Gln Pro Cys Leu Asp Leu Ile Glu Gly  
 46 110 115 120  
 47 Ser Pro Ile Ser Ile Cys Ser Leu Ile Asn Glu Glu Cys Arg Leu  
 48 125 130 135  
 49 Asn Arg Pro Ser Ser Ala Arg Gln Leu Gln Thr Arg Ile Glu Thr  
 50 140 145 150  
 51 Ala Leu Ala Gly Ser Pro Cys Leu Gly His Asn Lys Leu Ser Arg  
 52 155 160 165

ENTERED

## RAW SEQUENCE LISTING

DATE: 02/11/2003

PATENT APPLICATION: US/09/830,914C

TIME: 14:29:17

Input Set : A:\pf0621usn\_subseqlist.txt

Output Set: N:\CRF4\02112003\I830914C.raw

53	Glu	Pro	Ser	Phe	Ile	Val	Val	His	Tyr	Ala	Gly	Pro	Val	Arg	Tyr
54					170					175					180
55	His	Thr	Ala	Gly	Leu	Val	Glu	Lys	Asn	Lys	Asp	Pro	Ile	Pro	Pro
56					185					190					195
57	Glu	Leu	Thr	Arg	Leu	Leu	Gln	Gln	Ser	Gln	Asp	Pro	Leu	Leu	Met
58					200					205					210
59	Gly	Leu	Phe	Pro	Thr	Asn	Pro	Lys	Glu	Lys	Thr	Gln	Glu	Glu	Pro
60					215					220					225
61	Pro	Gly	Gln	Ser	Arg	Ala	Pro	Val	Leu	Thr	Val	Val	Ser	Lys	Phe
62					230					235					240
63	Lys	Ala	Ser	Leu	Glu	Gln	Leu	Leu	Gln	Val	Leu	His	Ser	Thr	Thr
64					245					250					255
65	Pro	His	Tyr	Ile	Arg	Cys	Ile	Lys	Pro	Asn	Ser	Gln	Gly	Gln	Ala
66					260					265					270
67	Gln	Thr	Phe	Leu	Gln	Glu	Glu	Val	Leu	Ser	Gln	Leu	Glu	Ala	Cys
68					275					280					285
69	Gly	Leu	Val	Glu	Thr	Ile	His	Ile	Ser	Ala	Ala	Gly	Phe	Pro	Ile
70					290					295					300
71	Arg	Val	Ser	His	Arg	Asn	Phe	Val	Glu	Arg	Tyr	Lys	Leu	Leu	Arg
72					305					310					315
73	Arg	Leu	His	Pro	Cys	Thr	Ser	Ser	Gly	Pro	Asp	Ser	Pro	Tyr	Pro
74					320					325					330
75	Ala	Lys	Gly	Leu	Pro	Glu	Trp	Cys	Pro	His	Ser	Glu	Glu	Ala	Thr
76					335					340					345
77	Leu	Glu	Pro	Leu	Ile	Gln	Asp	Ile	Leu	His	Thr	Leu	Pro	Val	Leu
78					350					355					360
79	Thr	Gln	Ala	Ala	Ala	Ile	Thr	Gly	Asp	Ser	Ala	Glu	Ala	Met	Pro
80					365					370					375
81	Ala	Pro	Met	His	Cys	Gly	Arg	Thr	Lys	Val	Phe	Met	Thr	Asp	Ser
82					380					385					390
83	Met	Leu	Glu	Leu	Leu	Glu	Cys	Gly	Arg	Ala	Arg	Val	Leu	Glu	Gln
84					395					400					405
85	Cys	Ala	Arg	Cys	Ile	Gln	Gly	Gly	Trp	Arg	Arg	His	Arg	His	Arg
86					410					415					420
87	Glu	Gln	Glu	Arg	Gln	Trp	Arg	Ala	Val	Met	Leu	Ile	Gln	Ala	Ala
88					425					430					435
89	Ile	Arg	Ser	Trp	Leu	Thr	Arg	Lys	His	Ile	Gln	Arg	Leu	His	Ala
90					440					445					450
91	Ala	Ala	Thr	Val	Ile	Lys	Arg	Ala	Trp	Gln	Lys	Trp	Arg	Ile	Arg
92					455					460					465
93	Met	Ala	Cys	Leu	Ala	Ala	Lys	Glu	Leu	Asp	Gly	Val	Glu	Glu	Lys
94					470					475					480
95	His	Phe	Ser	Gln	Ala	Pro	Cys	Ser	Leu	Ser	Thr	Ser	Pro	Leu	Gln
96					485					490					495
97	Thr	Arg	Leu	Leu	Glu	Ala	Ile	Ile	Arg	Leu	Trp	Pro	Leu	Gly	Leu
98					500					505					510
99	Val	Leu	Ala	Asn	Thr	Ala	Met	Gly	Val	Gly	Ser	Phe	Gln	Arg	Lys
100					515					520					525
101	Leu	Val	Val	Trp	Ala	Cys	Leu	Gln	Leu	Pro	Arg	Gly	Ser	Pro	Ser

## RAW SEQUENCE LISTING

DATE: 02/11/2003

PATENT APPLICATION: US/09/830,914C

TIME: 14:29:17

Input Set : A:\pf0621usn\_subseqlist.txt

Output Set: N:\CRF4\02112003\I830914C.raw

```

102          530          535          540
103 Ser Tyr Thr Val Gln Thr Ala Gln Asp Gln Ala Gly Val Thr Ser
104          545          550          555
105 Ile Arg Ala Leu Pro Gln Gly Ser Ile Lys Phe His Cys Arg Lys
106          560          565          570
107 Ser Pro Leu Arg Tyr Ala Asp Ile Cys Pro Glu Pro Ser Pro Tyr
108          575          580          585
109 Ser Ile Thr Gly Phe Asn Gln Ile Leu Leu Glu Arg His Arg Leu
110          590          595          600
111 Ile His Val Thr Ser Ser Ala Phe Thr Gly Leu Gly
112          605          610
114 <210> SEQ ID NO: 2
115 <211> LENGTH: 2109
116 <212> TYPE: DNA
117 <213> ORGANISM: Homo sapiens
119 <220> FEATURE:
120 <221> NAME/KEY: misc_feature
121 <223> OTHER INFORMATION: Incyte ID No: 1929760CB1
123 <400> SEQUENCE: 2
124 tgatgctctg ggctgtcttc acacttcatt tgggtttcct gcttgctctg agctctacag 60
125 gggaatgggg tagagatggg agccaccttg ggtggagggg ggggaaggta tgttctgccc 120
126 accacaggtg tcatgctcac tcagcctgat gccagggctg ccaagtataa ggcatgggca 180
127 ggggccagc caccctgggt tccttggtcc cctattcccc atctgctccc tgcagtggcc 240
128 ccatgggttc tctgccatct tcccaggcct gctggatgtg tatggatttg aatcatttcc 300
129 tgacaacagt ctggaacagt tgtgcatcaa ctacgccaat gagaagctgc agcagcattt 360
130 tgtggctcac tacctaaggg cccagcagga ggaatacgca gttgagggcc tggagtggtc 420
131 attcatcaac taccaggaca accagccctg tttgatctc attgagggaa gccccatcag 480
132 catctgctcc ctcataaatg aggaatgccg cctcaatcga cccagcagcg cagccagct 540
133 ccagacacgc attgagactg ccctggcagg cagccctgc ctgggccaca ataagctcag 600
134 ccgggagccc agcttcattg tgggtcatta tgcggggcct gtgcggtacc acacagcagg 660
135 cctggtggag aagaacaagg accctatccc acctgagctg accaggctcc tgcagcaatc 720
136 ccaggacccc ctgctcatgg ggctgtttcc tactaaccac aaagagaaga cccaggagga 780
137 accccctggc cagagcaggg cccctgtgtt gaccgtggtg tccaagttca aggcctcact 840
138 ggagcagctt ctgcaggctc tacacagcac cacgccccac tacattcgct gcatcaagcc 900
139 caacagccag ggccaggcgc agacctttct ccaagaggag gtcttgagcc agctggaggc 960
140 ctgtggcctc gtggagacca tccatatcag tgctgctggc ttccccatcc gggctctctca 1020
141 ccgaaacttt gtagaacgat acaagttact aagaaggctt catccttgca catcctctgg 1080
142 ccccgacagc ccatatcctg ccaaagggtt cctgaatgg tgtccacaca gcgaggaagc 1140
143 cagccttgaa cctctcatcc aggacattct ccacactctg ccggtcctaa ctcaggcagc 1200
144 agccataact ggtgactcgg ctgaggccat gccagcccc atgcactgtg gcaggaccaa 1260
145 ggtgttctga actgactcta tctggagct tctggaatgt gggcgtgccc ggggtgctgga 1320
146 gcagtgtgcc cgctgcatcc aggggtggctg gaggcgacac cggcaccgag agcaggagcg 1380
147 gcagtggcgg gccgtcatgc tcatccaggc agccattcgt tcctggttaa ctcggaaca 1440
148 catccagagg ctgcatgcag ctgccacagt catcaagcgt gcatggcaga agtggagaat 1500
149 cagaatggcc tgccttgctg ctaaagagct ggatggtgtg gaagaaaaac acttctctca 1560
150 agctccctgt tccctgagca cctcgccgct gcagaccagg ctctggagg caataatccg 1620
151 cctctggccc ctgggactgg tcttgccaa tacggctatg ggtgtaggca gctttcagag 1680
152 gaaattagtg gtctgggctt gcctccagct cccaggggc agccccagta gctacactgt 1740
153 ccagacagca caagaccagg ctggtgtcac gtccatccga gcgctgcctc agggatcgat 1800

```

## RAW SEQUENCE LISTING

DATE: 02/11/2003

PATENT APPLICATION: US/09/830,914C

TIME: 14:29:17

Input Set : A:\pf0621usn\_subseqlist.txt

Output Set: N:\CRF4\02112003\I830914C.raw

```

154 aaagtttcac tgcagaaagt ctccactgcg gtatgctgac atctgccctg aaccttcacc 1860
155 ctacagcatt acaggcttta atcagattct gctggaaaga cacaggctga tccacgtgac 1920
156 ctcttctgcc ttacttgggc tggggtgac cttggtgcct ttgtttccac aaggcctttt 1980
157 cctgccccct gccttgccaa agacatttaa tcagcacaca gctgccagac tattcccaca 2040
158 gtgctccaaa tgcacatgaa caacagtgcg ggctccagcc ttcgaccag agccccgtgc 2100
159 ccagtgcgt                                     2109

```

161 &lt;210&gt; SEQ ID NO: 3

162 &lt;211&gt; LENGTH: 1839

163 &lt;212&gt; TYPE: PRT

164 &lt;213&gt; ORGANISM: Caenorhabditis elegans

166 &lt;220&gt; FEATURE:

167 &lt;221&gt; NAME/KEY: misc\_feature

168 &lt;223&gt; OTHER INFORMATION: GenBank ID No: g1279777

170 &lt;400&gt; SEQUENCE: 3

```

171 Met Phe Asn Tyr Ser Lys Ile Phe Gln Ile His Arg Ala Cys Ser
172   1           5           10           15
173 Pro Asn Arg Lys Lys Ile Gly Ser Ile Gln Tyr Gly Arg Arg Arg
174           20           25           30
175 His Ser Trp Gln Gly Pro Val Val Pro Ala Ala Lys Leu Gln Val
176           35           40           45
177 Leu Ile Lys Gly Val Arg Ile Trp His Arg His Pro Thr Leu Val
178           50           55           60
179 Trp Ile Gly Ala Thr Leu Glu Glu Asp Ile Thr Phe Gln Thr Arg
180           65           70           75
181 Asn Val Arg Ile Arg Leu Glu Asp Asp Thr Glu Val Glu Tyr Ala
182           80           85           90
183 Ile Lys Ser Leu Asp Gln Leu Pro Phe Leu Arg Asn Pro Ala Phe
184           95          100          105
185 Leu Val Gly Lys Asp Asp Leu Thr Leu Leu Ser Tyr Leu His Glu
186          110          115          120
187 Pro Ala Val Leu His Asn Leu Gln Val Arg Phe Val Lys Gly Ser
188          125          130          135
189 Ser Ile Tyr Thr Tyr Cys Gly Ile Val Leu Val Ala Ile Asn Pro
190          140          145          150
191 Tyr Ala Asp Cys Ser His Ile Tyr Gly Glu Glu Ile Ile Gln Val
192          155          160          165
193 Tyr Arg Gly Ala Gly Lys Ser Ala Arg Glu Met Asp Pro His Ile
194          170          175          180
195 Phe Ala Val Ala Glu Glu Ala His Phe Asp Met Gly Ala Phe Gly
196          185          190          195
197 Lys Ser Gln Ser Ile Ile Val Ser Gly Glu Ser Gly Ala Gly Lys
198          200          205          210
199 Thr Val Ser Ala Lys Phe Val Met Arg Tyr Leu Ala Ser Val Ala
200          215          220          225
201 Ala Ser Lys Thr Arg Asn Gly Gly Thr Thr Ser Ile Glu Ala Arg
202          230          235          240
203 Val Leu Ala Ser Asn Pro Ile Met Glu Ser Ile Gly Asn Ala Lys
204          245          250          255
205 Thr Ile Arg Asn Asp Asn Ser Ser Arg Phe Gly Lys Phe Ile Gln

```

## RAW SEQUENCE LISTING

DATE: 02/11/2003

PATENT APPLICATION: US/09/830,914C

TIME: 14:29:17

Input Set : A:\pf0621usn\_subseqlist.txt

Output Set: N:\CRF4\02112003\I830914C.raw

206		260		265		270
207	Ile Asn Phe Cys	Glu Arg Gly Arg Arg	Ile Val Gly Ala Glu	Met		
208		275		280		285
209	Lys Thr Tyr Leu	Leu Glu Lys Ser Arg	Leu Val Phe Gln Ala	Pro		
210		290		295		300
211	Gly Glu Arg Asn	Tyr His Ile Phe Tyr	Gln Leu Cys Ala Ala	Arg		
212		305		310		315
213	Asn His Gln Val	Leu Lys Asp Leu His	Leu Gly Pro Cys Glu	Ser		
214		320		325		330
215	Tyr Ser Tyr Leu	Thr Gln Gly Gly Asp	Ser Arg Ile Pro Gly	Val		
216		335		340		345
217	Asp Asp Lys Ala	Asp Phe Glu Ala Leu	Leu Lys Ala Leu Gln	Leu		
218		350		355		360
219	Leu Gly Phe Asp	Glu Lys Gln Met Ser	Asp Val Phe Arg Leu	Leu		
220		365		370		375
221	Ala Gly Leu Leu	Leu Leu Gly Asn Val	His Phe Glu Asn Gly	Glu		
222		380		385		390
223	Gly Ser Ser Ala	Val Ser Ala Ser Ser	Cys Gln Glu Ile Ser	Arg		
224		395		400		405
225	Leu Cys Arg Glu	Phe Trp Lys Ile Ser	Glu Ser Asp Leu Arg	Ile		
226		410		415		420
227	Trp Leu Thr Arg	Arg Glu Ile Arg Ala	Val Asn Glu Ile Val	Thr		
228		425		430		435
229	Lys Pro Leu Thr	Lys Asn Glu Ala Val	Arg Ser Arg Asp Ala	Leu		
230		440		445		450
231	Thr Lys Met Leu	Tyr Ser His Leu Phe	Gly Trp Leu Val Asp	Lys		
232		455		460		465
233	Ile Asn Glu Ala	Leu Asn Glu Lys Asp	Lys Leu Asp Gly Thr	Asn		
234		470		475		480
235	Gln Lys Lys Arg	Pro Asp Arg Phe Ile	Gly Val Leu Asp Ile	Tyr		
236		485		490		495
237	Gly Phe Glu Thr	Phe Asp Val Asn Ser	Phe Glu Gln Phe Ser	Ile		
238		500		505		510
239	Asn Tyr Ala Asn	Glu Lys Leu Gln Gln	Gln Phe Asn Gln His	Val		
240		515		520		525
241	Phe Lys Leu Glu	Gln Glu Glu Tyr Ile	Arg Glu Glu Ile Glu	Trp		
242		530		535		540
243	Val Arg Val Asp	Phe His Asp Asn Gln	Pro Ala Ile Asp Leu	Ile		
244		545		550		555
245	Glu Gly Pro Val	Gly Met Ile Asn Leu	Leu Asp Glu Gln Cys	Lys		
246		560		565		570
247	Arg Leu Asn Gly	Ser Asp Ala Asp Trp	Leu Ser Gln Leu Gln	Asn		
248		575		580		585
249	Ser Thr Glu Leu	Lys Arg Asn Pro Gln	Leu Ala Phe Pro Lys	Val		
250		590		595		600
251	Arg Ser Asn Asp	Phe Ile Val Arg His	Phe Ala Ala Asp Val	Thr		
252		605		610		615
253	Tyr Ser Thr Asp	Gly Phe Val Glu Lys	Asn Arg Asp Ala Ile	Gly		
254		620		625		630

**VERIFICATION SUMMARY**

DATE: 02/11/2003

PATENT APPLICATION: US/09/830,914C

TIME: 14:29:18

Input Set : A:\pf0621usn\_subseqlist.txt

Output Set: N:\CRF4\02112003\I830914C.raw

L:10 M:271 C: Current Filing Date differs, Replaced Current Filing Date